Package 'icpack'

July 2, 2024

Type Package

Date 2024-06-21

Title Survival Analysis of Interval-Censored Data

Version 0.1.0

Depends R (>= 4.1.0), survival (>= 3.1)

Imports methods, rlang, gridExtra, checkmate, matrixStats, dplyr, reshape2, ggplot2

Suggests testthat, flexsurv, lemon, knitr, markdown

Description Survival analysis of interval-censored data with proportional hazards, and an explicit smooth estimate of the baseline log-hazard with P-splines.

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 7.3.1

VignetteBuilder knitr

NeedsCompilation no

Author Hein Putter [aut, cre], Paul Eilers [aut]

Maintainer Hein Putter <h.putter@lumc.nl>

Repository CRAN

Date/Publication 2024-07-02 06:30:05 UTC

Contents

`bbase` *Compute a B-spline basis*

Description

Compute a B-spline basis

Usage

```
bbase(x, xl = min(x), xr = max(x), nseg = 10, deg = 3)
```

Arguments

`x`	The vector of values for which the basis is to be evaluated
`xl`	The left boundary of the domain
`xr`	The right boundary of the domain
`nseg`	The number of inter-knot segments on the domain
`deg`	The degree of the B-splines (2 means quadratic, 3 means cubic, and so on)

Value

A matrix containing the basis

Examples

```
x = runif(100)
B = bbase(x, 0, 1, 20, 3)
```

drugusers *Interval-censored drug users data*

Description

Interval-censored drug users data

Usage

```
data(drugusers)
```

Format

Data from a cohort of 940 injecting drug users attending a hospital detoxification unit in Barcelona, Spain. Time is months between initiation of intravenous drug use and HIV seroconversion. A dataframe with five columns:

`left` Last negative HIV test (0 if first HIV test was positive)

`right` First positive HIV test (`Inf` if last HIV test was negative)

`period` Period of initiation of drug use, factor with levels "1972-1980", "1981-1985", "1986-1991", "1992-1997"

`gender` Gender, factor with levels "male" and "female"

`age` Age at initiation of drug use (years)

References

Gomez G, Calle ML, Egea JM & Muga R (2000). Risk of HIV infection as a function of the duration of intravenous drug use: a non-parametric Bayesian approach. Stat Med; 19:2641–2656.

Estep *Perform the E-step in the EM algorithm*

Description

Perform the E-step in the EM algorithm

Usage

```
Estep(H, Ic, R1, dead)
```

Arguments

`H`	Hazards per individual (in columns)
`Ic`	Censoring interval per individual, coded as 0/1 (in columns)
`R1`	Left truncation interval per individual, coded as 0/1 (in columns)
`dead`	Boolean vector (TRUE is event, FALSE is right censored)

Value

A list with two matrices

Y	Expected probability of event per bin per subject
R	Expected probability of at risk per bin per subject

`fillplot` *Fills space between two lines in a graph*

Description

Taken from mstate

Usage

```
fillplot(x, y1, y2, col)
```

Arguments

x	Points on the x-axis
y1	First set of points on y-axis
y2	Second set of points on y-axis
col	The color to fill space with

Value

Nothing

`fitit` *Fit proportional hazard model with smooth baseline hazard and (optional) interval censoring*

Description

Fit proportional hazard model with smooth baseline hazard and (optional) interval censoring

Usage

```
fitit(
  Y,
  R,
  dead,
  X,
  B,
  Ic,
  R1,
  cbx,
  Pdiff,
  Pridge,
  lambda,
  nit = 50,
  tol = 1e-06,
  tollam = 0.01,
  update_lambda = FALSE,
  ic_update = TRUE,
  monitor = FALSE
)
```

Arguments

Y	Events (matrix, number of bins by subjects)
R	Risk sets (matrix, number of bins by subjects)
dead	(Boolean vector, TRUE if event, FALSE if right censored)
X	Covariates (matrix, number of covariates (+1) by subjects)
B	B-spline basis matrix
Ic	Censoring interval per individual, coded as 0/1 (in columns)
R1	Left truncation interval per individual, coded as 0/1 (in columns)
cbx	Vector of starting values
Pdiff	B-spline part of penalty matrix
Pridge	Ridge part of penalty matrix (for intercept)
lambda	Smoothing parameter (number)
nit	Maximum number of iterations (integer)
tol	Tolerance for final fit
tollam	Tolerance for switching to lambda update
update_lambda	Automatic update of lambda (Boolean)
ic_update	Update risk and event probabilities (Boolean)
monitor	Monitor convergence (Boolean)

Value

A list with items

cbx	Vector of
ll	Poisson GLM log-likelihood
lambda	Final tuning parameter
pen	Penalty part of penalized log-likelihood
ed	Effetive dimension of the baseline hazard
nit1	Number of iterations used in first phase
nit	Total number of iterations used (first plus second phase)
tollam	Tolerance used for switching to lambda update

---

get_input_icfit *Get and check input of icfit*

Description

Get and check input of icfit

Usage

```
get_input_icfit(formula, data, entry)
```

Arguments

formula	A formula object with response of the left of a ~ operator and terms on the right. The response must be a survival object as returned by the 'Surv' function, with type either right', 'counting' or 'interval2'
data	A data frame in which to interpret the variable names in the 'formula'
entry	When appropriate, a vector of entry (left truncation) times, or a string indicating the column name in 'data' containing entry times; only used if Surv object is of type 'interval2'

Value

A list with items

Ymat	Matrix (number of subjects x 3) containing entry, left and right hand of intervals
X	Matrix (number of subjects x number of covariates + 1) with design matrix of covariates

`icfit` *Fit a proportional hazards model with baseline hazard modeled by P-splines*

Description

Fit a proportional hazards model with baseline hazard modeled by P-splines

Usage

```
icfit(
  formula,
  data,
  entry,
  lambda = 10,
  nt = 100,
  tmax,
  nseg = 20,
  bdeg = 3,
  pord = 2,
  nit = 50,
  tol = 1e-06,
  tollam = 0.01,
  kappa = 1e-06,
  update_lambda = TRUE,
  ic_update = TRUE,
  monitor = FALSE
)
```

Arguments

`formula`	A formula object with response of the left of a ~ operator and covariate terms on the right. The response must be a survival object as returned by the 'Surv' function, with type either right', 'counting' or 'interval2'
`data`	A data frame in which to interpret the variable names in the 'formula'
`entry`	When appropriate, a vector of entry (left truncation) times, or a string indicating the column name in 'data' containing entry times; only used if Surv object is of type 'interval2'
`lambda`	Starting value of penalty tuning parameter
`nt`	The number of time bins
`tmax`	The end of time domain (default 1.01 times largest observation)
`nseg`	The number of B-spline segments
`bdeg`	The degree of the B-splines
`pord`	The order of the differences used in the penalty

nit	Maximum number of iterations (integer)
tol	Tolerance for final fit
tollam	Tolerance for switching to lambda update
kappa	Ridge parameter (number)
update_lambda	Automatic update of lambda (Boolean)
ic_update	Update risk and event probabilities (Boolean)
monitor	Monitor convergence (Boolean)

Value

An object of class 'icfit'

Examples

```
# Fit proportional hazards model to interval-censored data
icfit(Surv(left, right, type='interval2') ~ period + gender + age,
      data=drugusers)
# Fit proportional hazards model to right-censored data
icfit(Surv(time, d) ~ Diameter + FIGO + Karnofsky, data = Ova)
```

InfoMatrix *Compute the information matrix*

Description

Compute the information matrix

Usage

```
InfoMatrix(object, initres)
```

Arguments

object	Fit obtained from fitit
initres	Result from init

Details

Three information matrices are computed. One is Ifull which interprets the imputed R and Y data from object as actual observations. Iloss gives the loss of information due to imputation. The sum of both matrices is the true information matrix.

Value

A list with three items

`Itrue`	Total of `Ifull` and `Iloss`, true Fisher information matrix
`Ifull`	Full Fisher information matrix
`Iloss`	Loss of information due to intervals (missing event times)

`init` *Generate a discrete IC object*

Description

Generate a discrete IC object

Usage

```
init(Times, X, nt, tmax, nseg = 20, bdeg = 3, pord = 2, kappa = 1e-06)
```

Arguments

`Times`	The (possibly interval censored) survival data, in a matrix
`X`	The design matrix containing covariates
`nt`	The number of bins for discretization
`tmax`	The end of time domain (default 1.01 times largest observation)
`nseg`	The number of B-spline segments
`bdeg`	The degree of the B-splines
`pord`	The order of the differences used in the penalty
`kappa`	Ridge parameter (number)

Value

A list with items

`data`	List containing the original data as well as the binned data
`bins`	List with information on bins used
`basis`	List containing the B-spline matrix
`start`	List containing information on starting values
`penalty`	List containing Pdiff and Pridge
`control`	List with information on control of B-spline basis

Mstep *Function for fitting proportioal hazard model with baseline hazard*

Description

Function for fitting proportioal hazard model with baseline hazard

Usage

```
Mstep(Y, R, X, B, Pen, lambda, cbx)
```

Arguments

Y	Expected events (matrix)
R	Expected risk sets (matrix)
X	Covariates (matrix)
B	B-spline basis
Pen	Penalty matrix
lambda	Smoothing parameter (number)
cbx	Current coefficient estimates

Value

An object with fields: H = hazards (matrix), cbx = coefficient estimates (vector), lambda = proposal for new lambda, ed = effective dimension, G = G matrix, ll = log-likelihood, pen = penalized part of log-likelihood, Mpen = penalized M matrix

Ova *Ovarian cancer data*

Description

Ovarian cancer data

Usage

```
data(Ova)
```

Format

A dataframe with five columns:

Diameter
FIGO
Karnofsky
time
d death

Source

tba

`plot.icfit` *Plot method for an object of class 'icfit'*

Description

Plot method for an object of class 'icfit'

Usage

```
## S3 method for class 'icfit'
plot(
  x,
  type = c("hazard", "cumhazard", "survival", "probability"),
  conf.int = TRUE,
  ylim = NULL,
  title = NULL,
  xlab = NULL,
  ylab = NULL,
  fill = TRUE,
  fillcol = "lightgrey",
  ...
)
```

Arguments

`x`	The object of class 'icfit' to be plotted
`type`	Type of plot. Accepted choices: 'hazard' (default), 'cumhazard', 'survival' or 'cumprob'
`conf.int`	If 'TRUE' a 100*(1 - alpha) percent confidence interval is plotted
`ylim`	The y-limits for the plot
`title`	Optional title string
`xlab`	Text for x-label
`ylab`	Text for y-label
`fill`	Fill area between lower and upper
`fillcol`	The color for filling (default 'lightgrey')
`...`	Other arguments to plot (except 'type', which is set to 'l')

Value

A ggplot grob, containing the plot. Use `print()` or `plot()` to show it Multiple objects can be combined by using functions in the package `gridExtra`.

Examples

```
icf <- icfit(Surv(left, right, type='interval2') ~ period + gender + age,
             data = drugusers)
plot(icf)
```

`plot.predict.icfit` *Plot method for an object of class 'predict.icfit'*

Description

Plot method for an object of class 'predict.icfit'

Usage

```
## S3 method for class 'predict.icfit'
plot(
  x,
  type = c("hazard", "cumhazard", "survival", "probability"),
  conf.int = TRUE,
  fill = TRUE,
  fillcol = "lightgrey",
  ylim = NULL,
  title = NULL,
  xlab = NULL,
  ylab = NULL,
  selection = NULL,
  nrow = NULL,
  ncol = NULL,
  do_plot = TRUE,
  ...
)
```

Arguments

`x`	The object of class 'predict.icfit' to be plotted
`type`	Type of plot. Accepted choices: 'hazard' (default), 'cumhazard', 'survival' or 'probability'
`conf.int`	If 'TRUE' a 100*(1 - alpha) percent confidence interval is plotted
`fill`	Fill area between lower and upper
`fillcol`	The color for filling (default 'lightgrey')
`ylim`	The y-limits for the plot
`title`	Optional title string, or, if x is a list, obtained from 'predict.icfit' using 'newdata', a vector of title strings

xlab	Text for x-label
ylab	Text for y-label
selection	If x is a list, obtained from 'predict.icfit' using 'newdata', then a vector containing the subset of list elements to be plotted, default is to plot all elements of the list
nrow	If x is a list, obtained from 'predict.icfit' using 'newdata', then a number specifying the number of rows to plot; default the square root of the number of list elements to be plotted
ncol	If x is a list, obtained from 'predict.icfit' using 'newdata', then a number specifying the number of columns to plot; default the square root of the number of list elements to be plotted
do_plot	Boolean indicating whether or not to actually plot (default is TRUE)
...	other graphical parameters to be passed on

Value

A ggplot grob, containing the plot. Use `print()` or `plot()` to show it Multiple objects can be combined by using functions in the package `gridExtra`.

Examples

```
icf <- icfit(Surv(left, right, type='interval2') ~ period + gender + age, data=drugusers)
pred_icf <- predict(icf)
plot(pred_icf)
library(ggplot2)
plot(icf) + xlim(0, 200) + ylim(0, 0.05)
ndata <- drugusers[1:4, ]
pred_nd_icf <- predict(icf, newdata=ndata)
plot(pred_nd_icf) # plot all four
plot(pred_nd_icf[[2]]) # plot only the second
plot(pred_nd_icf, type = "cumhazard") # plot four cumulative hazard curves
plot(pred_nd_icf[[3]], type = "prob", ylim = c(0, 1)) # plot probability curve for nr 3
plot(pred_nd_icf[[4]], type = "surv", ylim = c(0, 1)) # plot survival curve for nr 4
```

---

predict.icfit *Predict method for an object of class 'icfit'*

Description

Predict method for an object of class 'icfit'

Usage

```
## S3 method for class 'icfit'
predict(object, newdata, nstep = 500, alpha = 0.05, ...)
```

Arguments

object	The object of class 'icfit' for which a prediction is to be made
newdata	A data frame containing covariate information for a new subject
nstep	Number of time steps used for calculating cumulative hazards (default is 500)
alpha	The alpha level for the (1-alpha)*100 percent confidence interval
...	Any other arguments

Value

An object of class 'predict.icfit', which is a data frame with time points and hazard, cumulative hazard and survival at those time points, along with standard errors and pointwise lower and upper confidence bounds, or a list of such data frames for each subject represented in 'newdata'

Examples

```
icf <- icfit(Surv(left, right, type='interval2') ~ period + gender + age, data=drugusers)
pred_icf <- predict(icf)
head(pred_icf)
ndata <- drugusers[1:4, ]
pred_nd_icf <- predict(icf, newdata=ndata)
lapply(pred_nd_icf, head)
```

---

print.icfit *Print method for an object of class 'icfit'*

Description

Print method for an object of class 'icfit'

Usage

```
## S3 method for class 'icfit'
print(x, digits = max(1L, getOption("digits") - 3L), alpha = 0.05, ...)
```

Arguments

x	The object of class 'icfit' to be printed
digits	Number of digits to be printed
alpha	Alpha level to be used of confidence interval ((1-alpha) * 100 percent)
...	Further arguments to print

Value

No return value

`rasterplot` *Plot probabilities as a raster‘*

Description

Plot probabilities as a raster‘

Usage

```
rasterplot(
  icf,
  type = c("both", "R", "Y"),
  sel = NULL,
  label = NULL,
  show_label = FALSE,
  pow = 0.2,
  order = TRUE,
  do_plot = TRUE
)
```

Arguments

`icf`	an object of class 'icfit'
`type`	a string giving the type of the plot. Accepted choices: 'R' (risk probabilities) and 'Y' (event probabilities)
`sel`	a vector of integers for selection of subject (rows of the matrix)
`label`	character vector containing labels for the individuals to be plotted in selection
`show_label`	Boolean, whether or not to show the labels
`pow`	a number, giving he power to which the probabilities will be raised, to improve the clarity of the plot
`order`	Boolean, default (TRUE) is to order according to first positive in Y, then first zero in Y, then first zero in R; if FALSE order of occurrence in data is used
`do_plot`	Boolean, default (TRUE) shows the plot, if FALSE object is returned but not plotted

Value

a ggplot object (Grob)

Examples

```
icf <- icfit(Surv(left, right, type='interval2') ~ period + gender + age,
  data=drugusers)
rasterplot(icf)
rasterplot(icf, type = 'R')
rasterplot(icf, type = 'Y')
```

```
rasterplot(icf, pow = 0.05) # very small power basically shows 0/1
sel <- c(
  11, 18,  # right-censored, event in (L, \infty)
  1:2,     # event in (0, R)
  115, 133 # event in (L, R)
)
rasterplot(icf, sel = sel)
rasterplot(icf, sel = sel, label = c("e", "p", "g", "c", "m", "n"), show_label = TRUE)
rasterplot(icf, sel = sel, label = c("e", "p", "g", "c", "m", "n"), show_label = TRUE,
  type = 'Y')
```

---

summary.icfit *Summary method for an object of class 'icfit'*

Description

Summary method for an object of class 'icfit'

Usage

```
## S3 method for class 'icfit'
summary(
  object,
  lvl = 1,
  digits = max(1L, getOption("digits") - 3L),
  alpha = 0.05,
  ...
)
```

Arguments

object	Object of class 'icfit'
lvl	Describes the level of output
digits	Number of digits to be printed
alpha	Alpha level to be used of confidence interval ((1-alpha) * 100 percent)
...	Other arguments to summary

Value

None (invisible NULL)

Examples

```
icf <- icfit(Surv(left, right, type='interval2') ~ period + gender + age, data=drugusers)
summary(icf)
summary(icf, lvl=0) # same as print(icf)
summary(icf, lvl=1) # extra information on iterations and computation time
```

---

summary.predict.icfit *Summary method for an object of class 'predict.icfit'*

Description

Summary method for an object of class 'predict.icfit'

Usage

```
## S3 method for class 'predict.icfit'
summary(object, times, ...)
```

Arguments

object	Object of class 'predict.icfit'
times	The time points at which to summarize the predicted hazards, cumulative hazards and survival probabilities, with associated standard errors and confidence intervals
...	Other arguments to plot

Value

A data frame (if object was a data frame) or a list of data frames (if object was a list of data frames) with hazards etc linearly interpolated between the time points used in the predict function

Examples

```
icf <- icfit(Surv(left, right, type='interval2') ~ period + gender + age,
      data=drugusers)
pred_icf <- predict(icf)
summary(pred_icf, times=c(0, 30, 183, 365))
```

Index